NeuroImage 218 (2020) 116932

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Amygdalar nuclei and hippocampal subfields on MRI: Test-retest reliability of automated volumetry across different MRI sites and vendors

Giulia Quattrini [a,*], Michela Pievani [a], Jorge Jovicich [b], Marco Aiello [c], Núria Bargalló [d], Frederik Barkhof [e,f], David Bartres-Faz [g], Alberto Beltramello [h], Francesca B. Pizzini [i], Olivier Blin [j], Regis Bordet [k], Massimo Caulo [l], Manos Constantinides [m], Mira Didic [n,o], Antonios Drevelegas [m], Antonio Ferretti [l], Ute Fiedler [p], Piero Floridi [q], Hélène Gros-Dagnac [r], Tilman Hensch [s], Karl-Titus Hoffmann [t], Joost P. Kuijer [e], Renaud Lopes [u], Camillo Marra [v], Bernhard W. Müller [w], Flavio Nobili [x,y], Lucilla Parnetti [z], Pierre Payoux [r], Agnese Picco [x], Jean-Philippe Ranjeva [aa], Luca Roccatagliata [y,ab], Paolo M. Rossini [ac], Marco Salvatore [c], Peter Schonknecht [s], Björn H. Schott [ad,ae,am], Julien Sein [aa], Andrea Soricelli [c], Roberto Tarducci [af], Magda Tsolaki [ag], Pieter J. Visser [ah,ai], Jens Wiltfang [ad,al,am], Jill C. Richardson [aj], Giovanni B. Frisoni [a,ak], Moira Marizzoni [a], on behalf of PharmaCog Consortium

[a] Laboratory of Alzheimer's Neuroimaging and Epidemiology (LANE), IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy
[b] Center for Mind Brain Sciences, University of Trento, Trento, Italy
[c] IRCCS SDN, Napoli, Italy
[d] Department of Neuroradiology and Image Research Platform, Hospital Clínic de Barcelona, IDIBAPS, Barcelona, Spain
[e] Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, the Netherlands
[f] Queen Square Institute of Neurology and Centre for Medical Image Computing, University College London, UK
[g] Department of Medicine and Health Sciences, Faculty of Medicine, Universitat de Barcelona and IDIBAPS, Barcelona, Spain
[h] Department of Radiology, IRCCS "Sacro Cuore-Don Calabria", Negrar, Verona, Italy
[i] Radiology, Department of Diagnostic and Public Health, University of Verona, Verona, Italy
[j] Aix-Marseille University, UMR-INSERM 1106, Service de Pharmacologie Clinique, APHM, Marseille, France
[k] Aix-Marseille Université, INSERM U 1106, 13005, Marseille, France
[l] University "G. d'Annunzio" of Chieti, Chieti, Italy
[m] Interbalkan Medical Centre of Thessaloniki, Thessaloniki, Greece
[n] Aix-Marseille Université, Inserm, Institut de Neurosciences des Systèmes (INS) UMR_S 1106, 13005, Marseille, France
[o] APHM, Timone, Service de Neurologie et Neuropsychologie, Hôpital Timone Adultes, Marseille, France
[p] Institutes and Clinics of the University Duisburg-Essen, Essen, Germany
[q] Perugia General Hospital, Neuroradiology Unit, Perugia, Italy
[r] ToNIC, Toulouse NeuroImaging Center, Université de Toulouse, Inserm, UPS, France
[s] Department of Psychiatry and Psychotherapy, University of Leipzig Medical Center, Leipzig, Germany
[t] Department of Neuroradiology, University Hospital Leipzig, Leipzig, Germany
[u] INSERM U1171, Neuroradiology Department, University Hospital, Lille, France
[v] Catholic University, Fondazione Policlinico A. Gemelli, IRCCS, Rome, Italy
[w] LVR-Hospital Essen, Department for Psychiatry and Psychotherapy, Faculty of Medicine, University of Duisburg-Essen, Germany
[x] Department of Neuroscience (DINOGMI), University of Genoa, Genoa, Italy
[y] IRCCS, Ospedale Policlinico San Martino, Genova, Italy
[z] Section of Neurology, Department of Medicine, University of Perugia, Perugia, Italy
[aa] CRMBM-CEMEREM, UMR 7339, Aix-Marseille University, CNRS, Marseille, France
[ab] Department of Health Science (DISSAL), University of Genoa, Genoa, Italy
[ac] Dept. Neuroscience & Rehabilitation, IRCCS San Raffaele-Pisana, Rome, Italy
[ad] Department of Psychiatry and Psychotherapy, University Medical Center Goettingen (UMG), Göttingen, Germany
[ae] Leibniz Institute for Neurobiology, Magdeburg, Germany
[af] Perugia General Hospital, Medical Physics Unit, Perugia, Italy
[ag] Aristotle University of Thessaloniki, Thessaloniki, Greece
[ah] Department of Neurology, Alzheimer Centre, VU Medical Centre, Amsterdam, Netherlands

* Corresponding author. IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Via Pilastroni 4, 25125, Brescia, Italy.

https://doi.org/10.1016/j.neuroimage.2020.116932

Available online 13 May 2020

[ai] Maastricht University, Maastricht, Netherlands
[aj] Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Gunnels Wood Road, Stevenage, United Kingdom
[ak] Memory Clinic and LANVIE-Laboratory of Neuroimaging of Aging, Hospitals and University of Geneva, Geneva, Switzerland
[al] Neurosciences and Signaling Group, Institute of Biomedicine (iBiMED), Department of Medical Sciences, University of Aveiro, Aveiro, Portugal
[am] German Center for Neurodegenerative Diseases (DZNE), Goettingen, Germany

ARTICLE INFO

Keywords:
Amygdalar nuclei
Hippocampal subfields
Multicenter MRI study
FreeSurfer
Reliability analysis

ABSTRACT

Background: The amygdala and the hippocampus are two limbic structures that play a critical role in cognition and behavior, however their manual segmentation and that of their smaller nuclei/subfields in multicenter datasets is time consuming and difficult due to the low contrast of standard MRI. Here, we assessed the reliability of the automated segmentation of amygdalar nuclei and hippocampal subfields across sites and vendors using FreeSurfer in two independent cohorts of older and younger healthy adults.
Methods: Sixty-five healthy older (cohort 1) and 68 younger subjects (cohort 2), from the PharmaCog and CoRR consortia, underwent repeated 3D-T1 MRI (interval 1–90 days). Segmentation was performed using FreeSurfer v6.0. Reliability was assessed using volume reproducibility error (ε) and spatial overlapping coefficient (DICE) between test and retest session.
Results: Significant MRI site and vendor effects ($p < .05$) were found in a few subfields/nuclei for the ε, while extensive effects were found for the DICE score of most subfields/nuclei. Reliability was strongly influenced by volume, as ε correlated negatively and DICE correlated positively with volume size of structures (absolute value of Spearman's r correlations >0.43, $p < 1.39E-36$). In particular, volumes larger than 200 mm^3 (for amygdalar nuclei) and 300 mm^3 (for hippocampal subfields, except for molecular layer) had the best test-retest reproducibility (ε < 5% and DICE > 0.80).
Conclusion: Our results support the use of volumetric measures of larger amygdalar nuclei and hippocampal subfields in multisite MRI studies. These measures could be useful for disease tracking and assessment of efficacy in drug trials.

1. Introduction

The amygdala and the hippocampus are two limbic structures that play a critical role in cognitive functions and behavior, such as memory, learning, and emotional processes (Catani et al., 2013; Rolls, 2015). Both structures encompass several subregions (i.e., nuclei and subfields, for the amygdala and the hippocampus, respectively), defined by the specific cytoarchitectonics and connections. Several amygdalar parcellation strategies have been proposed based on histological and/or animal data (Amunts et al., 2005; Heimer et al., 1999; Sah et al., 2003) and a recent study delineated amygdala subregions based on *in vivo* MRI (Tyszka and Pauli, 2016). One common strategy groups the amygdalar nuclei into the basolateral complex (lateral, basal, and accessory basal nuclei), the cortical-like group (cortical nucleus, nucleus of the lateral olfactory tract, bed nucleus of the accessory olfactory tract, and peri-amygdaloid cortex), the centromedial complex (central and medial nuclei, and amygdaloid part of the bed nucleus of the stria terminalis), the anterior amygdaloid area, the intercalated nuclei, and the amygdalohippocampal area (Sah et al., 2003). The basolateral complex is considered as the input region of the amygdala, receiving sensory information from cortex and thalamus, and is related to fear conditioning, adaptive and goal-directed behaviors (for a review on human and animal studies, see Wassum and Izquierdo, 2015). The basolateral complex is mainly composed by excitatory glutamatergic neurons, showing a pyramidal morphology, and by a lower number of gamma-aminobutyric acidergic (GABAergic) inhibitors, defining an intrinsic circuitry for the antagonistic control of emotional behaviors and memories (Braak and Braak, 1983; Kim et al., 2016; Krabbe et al., 2018). Nuclei of the cortical-like group, which mainly show a layered structure, receive stimuli from the olfactory bulb and are supposed to be involved in stimuli reinforce processes (due to projections to the striatum) and in memory processing (due to connections with the hippocampal and parahippocampal areas) (Feher, 2017; Kemppainen et al., 2002; Sah et al., 2003; Ubeda-Bañon et al., 2007). Finally, the centromedial complex, the striatum-like output region of the amygdala, mainly projects to the brainstem and the hypothalamus for the regulation of fear- and stress-related reactions (Krabbe et al., 2018; Yang and Wang, 2017). In particular, the centromedial complex, which includes GABAergic neurons, receives sensory information from the basolateral complex, and sends inhibitory projections for the regulations of fear and anxiety responses (Babaev et al., 2018; van den Burg and Stoop, 2019).

The hippocampal formation consists of distinct subfields along the transverse axis: the cornu ammonis (CA) 1, 3, 4, including pyramid neurons; the dentate gyrus, showing a granular cell layer and a thick molecular layer; the subicular complex, also mainly composed by pyramid neurons, and discernable in subiculum, prosubiculum, presubiculum, postsubiculum, and parasubiculum; and the hippocampus-amygdala transition area, mainly formed by densely packed cells (Ding, 2013; Ding and Van Hoesen, 2015). The intrinsic connectivity is described by trisynaptic loop of the hippocampus: from the entorhinal cortex (the major cortical input) to the dentate gyrus, then to the CA3, and finally to the CA1 (output region) (Knierim, 2015). The other core pathway includes the direct connection between the entorhinal cortex and the subiculum (Knierim, 2015). Functional studies evidenced three distinct subregions along the longitudinal axis, showing specialized connectivity patterns: i) the head (anterior segment), which connects to the prefrontal cortex; ii) the body (intermediate segment), linking to the posterior cingulate cortex; and iii) the tail (posterior segment), related to the thalamus (Zarei et al., 2013; Zhong et al., 2019). The different connectivity pattern of the anterior-posterior axis also reflect a functional specialization: while the head seems to be associated to emotional, motivation/reward, and memory encoding processes, the body and tail are probably related to spatial navigation and memory retrieval (for reviews on human and animal studies, see Poppenk et al., 2013 and Strange et al., 2014).

Within the limbic system, the amygdalar sub-circuit includes the orbitofrontal cortex and the anterior cingulate cortex, and is mainly related to emotional processes (particularly fear-related behavior, such as fear conditioning and recognition of fearful stimuli) and reward-related mechanisms (Benarroch, 2015; Janak and Tye, 2015; Murray et al., 2014; Rolls, 2015). On the other hand, the hippocampal sub-circuit participates to episodic memories formation, spatial learning and navigation, due to connections with the posterior cingulate cortex and the fornix-mammillary body-anterior thalamus-posterior cingulate circuit (Howard and Eichenbaum, 2015; Janak and Tye, 2015; Knierim, 2015; Rolls, 2015). Moreover, human hippocampus is also thought to be involved in the behavioral regulation according to the social context

(Montagrin et al., 2018).

The amygdala and the hippocampus are of great interest in the psychiatric and aging research fields, due to their involvement in several mental/cognitive disorders (Benson et al., 2014; Ganzola et al., 2014; Rossi et al., 2012) and neurodegenerative diseases (Bang et al., 2015; Bouchard et al., 2008; Frisoni et al., 2008; Prestia et al., 2011; Cavedo et al., 2011).

Structural neuroimaging techniques can be used to measure amygdala and hippocampal structures *in vivo*. Although manual segmentation is still considered as the gold standard for volumetric measurement, it is inherently subjective and suffers from high inter-rater variability. Moreover, the resolution of standard MRI acquisition protocols (0.75–1 mm^3 voxel size) does not allow to clearly visualize the boundaries of small structures such as nuclei of the amygdala and hippocampal subfields as enabled by histology or high field (7T) MRI (Saygin et al., 2017). Finally, manual segmentation is time-consuming and thus not suitable for studies involving a large number of subjects. This is particularly relevant for multicenter studies, which are increasingly being carried out worldwide and typically include hundreds or thousands of scans. To overcome the shortcomings of manual segmentation, several automated and semi-automated algorithms have been developed (for a review, see Despotović et al., 2015; Helms, 2016). Importantly, multicenter reliability studies using morphometric measures must consider differences in MRI hardware and acquisition protocols across sites in order to control for site variability, which could affect the interpretation of results (Han et al., 2006; Jovicich et al., 2009; Wonderlick et al., 2009).

Among the available automated tools, FreeSurfer (Fischl, 2012; https://surfer.nmr.mgh.harvard.edu/) is one of the most widely used. Previous studies showed a general consistency between the automated and the manual segmentation of the hippocampus and the amygdala using the FreeSurfer software suite, both in healthy and clinical populations (Tae et al., 2008; Morey et al., 2009, 2010; Dewey et al., 2010; Mulder et al., 2014; Iscan et al., 2015; Grimm et al., 2015). Recently, a new atlas for the segmentation of the amygdalar nuclei and the hippocampal subfields (Iglesias et al., 2015; Saygin et al., 2017) has been released as part of FreeSurfer v6.0. This feature allows the simultaneous segmentation of both structures, thereby avoiding structural overlap between them, which represents a major problem in neuroimaging studies involving these regions (Amunts et al., 2005; Saygin et al., 2017). The atlas was built using an *ex vivo*, high-resolution (100–150 μm isotropic) imaging protocol acquired at 7T magnetic field strength, which allows for a higher resolution and signal-to-noise ratio than *in vivo* procedures, making boundaries clearly visible (Iglesias et al., 2015; Saygin et al., 2017). Finally, compared to approaches that estimate probability maps in a reference space after warping of the T1 image to a standard template (Amunts et al., 2005; Tyszka and Pauli, 2016) the tool extracts individual segmentation in subject space, providing a better spatial sensitivity as compared to standard template space (Iglesias et al., 2015; Saygin et al., 2017). To the best of our knowledge, only two studies(Brown et al., 2020; Worker et al., 2018) has assessed the test-retest reliability of this new FreeSurfer tool so far. However, these reproducibility analyses did not consider the amygdala and did not test the effect of averaging two within-session T1 images, which has been shown to improve both volume and spatial reproducibility of hippocampal subfields segmentation in the previous FreeSurfer version (Marizzoni et al., 2015).

The aim of the present study is to assess the test-retest reproducibility across different MRI scanners of both the cross-sectional and longitudinal pipelines of amygdalar nuclei and hippocampal subfields in two independent cohorts, old and young healthy subjects.

2. Materials and methods

The data that support the findings of this study are openly available in at https://neugrid4you.eu/datasets and http://fcon_1000.projects.nitrc.org/indi/CoRR/html/.

2.1. Participants

2.1.1. Cohort 1- older volunteers

This study follows procedures similar to those described in previous reliability studies from the PharmaCog project (https://www.alzheimer-europe.org/Research/PharmaCog) on morphometric measures (Jovicich et al., 2013; Marizzoni et al., 2015). Briefly, 65 healthy old subjects from 13 clinical MRI sites across Europe were included: Italy (Chieti, Genoa, Naples, Perugia, and Verona), France (Lille, Marseille, and Toulouse), Germany (Essen and Leipzig), Spain (Barcelona), Netherlands (Amsterdam), and Greece (Thessaloniki). Inclusion and exclusion criteria have been described previously (Jovicich et al., 2014). Each site contributed 10 3D T1-weighted MRI scans obtained from local volunteers within an age range of 50–80 years (Table 1). All images were acquired at field strength of 3T. Each subject was scanned twice within a period ranging from 7 to 60 days, to minimize biological changes that could affect the reliability of the measures (Table 1). The study received the authorization from the local ethics committees or institutional review boards of the participating institutions, and all subjects provided written informed consent in accordance with the Declaration of Helsinki.

2.1.2. Cohort 2- younger volunteers

The cohort of younger volunteers consisted of 68 healthy subjects (age range 18–50 years) with a test-retest period ranging from 1 to 90 days (Table 1). MRI data, derived from American (Roanoke, New York, and Madison) and Chinese (Beijing) imaging groups, were publicly available and downloaded from the Consortium for Reliability and Reproducibility (CoRR; http://fcon_1000.projects.nitrc.org/indi/CoRR/html/; Zuo et al., 2014). Data collection and sharing were conducted with the approval of their respective local ethical committee or institutional review board (Zuo et al., 2014).

2.2. MRI acquisition

2.2.1. Cohort 1- older volunteers

The general harmonized acquisition protocol was previously described (Jovicich et al., 2013), and parameters are reported in Table 1. MRI scanners from different manufacturers were used by participating sites: Siemens (Allegra, TrioTim, Verio, Skyra and Biograph mMR), GE (HDxt and Discovery MR750), and Philips (Achieva). For each test and retest session, two anatomical 3D T1-weighted MPRAGE sequences were acquired, one at the beginning and one at the end of the MRI session. Only vendor-provided sequences were used.

2.2.2. Cohort 2- younger volunteers

MRI acquisition protocols for each MRI site are summarized in Table 1. MRI systems from two manufacturers were used (Siemens and GE), with three different models (TrioTim and Allegra from Siemens, Discovery from GE). For each subject, two separate 3D T1-weighted scans were obtained at 3T, one for the test session and one for the retest session respectively.

2.3. MRI processing

All MRI scans included in this study underwent a visual quality check before processing, according to a previously published rating system (Backhausen et al., 2016), and were processed on the same operating system and workstation using both the cross-sectional (Fischl et al., 2002) and the longitudinal (Reuter et al., 2012) standard pipelines implemented in FreeSurfer v6.0. FreeSurfer (Fischl et al., 1999; Dale et al., 1999) is a software suite designed to analyze structural and functional neuroimaging data.

2.3.1. Cross-sectional and longitudinal pipelines

The cross-sectional pipeline involves the following steps: i) motion correction and averaging (in case of multiple T1 weighted images); ii)

Table 1
Summary of the demographic and MRI features of the study cohorts by MRI sites.

OLDER COHORT (N = 65)													YOUNGER COHORT (N = 68)				
Site	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
MRI site	Verona	Barcelona	Leipzig	Marseille	Essen	Naples	Genoa	Thessaloniki	Amsterdam	Lille	Toulouse	Chieti	Perugia	New York	Roanoke	Beijing	Madison
N	5	5	5	5	5	5	5	5	5	5	5	5	5	20	5	34	9
Age (min–max), years	52–78	72–78	60–66	58–78	51–54	56–65	56–61	51–62	52–73	55–68	53–65	65–76	50–74	20–43	20–36	18–25	21–32
Sex (female/N)	2/5	5/5	3/5	4/5	2/5	2/5	2/5	3/5	3/5	3/5	3/5	5/5	3/5	8/20	2/5	23/34	3/9
Test-retest interval (min–max), days	7–49	7–16	7–35	7–56	7–38	7–34	9–48	23–41	7–21	7–22	8–26	8–21	7–14	1–85	51–89	1–24	56–81
MRI vendor	Siemens						GE			Philips				Siemens			GE
MRI scanner	Allegra	TrioTim		Verio	Skyra	Biograph mMR	HDxt		Discovery	Achieva				Allegra	TrioTim		Discovery
Voxel size	1 × 1 × 1													1.3 × 1 × 1.3	1.3 × 1 × 1		1 × 1 × 1
TE (ms)	2.83	2.98	2.98	2.98	2.03	2.96	2.86	2.98	2.92	3.16	3.16	3.10	3.10	3.25	3.02	2.51	3.18
MPRAGE acquisition time (min:sec)	9:50	5:12	5:12	5:12	5:12	5:03	4:43	4:40	4:14	6:50	6:50	6:08	6:08	8:07	4:38	5:53	7:30

Abbreviations: min, minimum value; max, maximum value; TE, echo time; min, minutes; sec, seconds.

non-uniformity correction and skull-stripping; iii) linear and non-linear Talairach transformation from native to MNI305 standard space; iv) intensity normalization, white matter and subcortical gray-matter (including hippocampus and amygdala) segmentation; v) gray/white matter boundary tessellation, topology correction, surface deformation following intensity gradients to derive white and pial surfaces; vi) surfaces inflation and registration to a spherical atlas based on individual patterns of cortical folding; vii) cortex parcellation according to gyri and sulci. Both the cortical and subcortical parcellations are provided by FreeSurfer in subject's native space. The cross-sectional pipeline was applied to both time-points for each subject.

The longitudinal stream is applied after the cross-sectional standard processing of T1 scans at the two time-points. This pipeline creates a template for each subject, using information from both time-points, to estimate the average subject anatomy and to reduce the variability (Reuter et al., 2012). Individual time-points are then automatically resampled to the template space using a robust, inverse consistent registration. The above processing steps (skull stripping, computation of Talairach transformations,non-linear registration to the atlas, spherical surface mapping and parcellation) are then initialized using common information from the within-subject template.

For the older cohort, the within-session MRI images were automatically averaged prior to the processing and analyzed using the longitudinal pipeline. While the co-registration between the test and the retest time-points was automatically performed as part of the longitudinal pipeline, the FMRIB Software Library as implemented in FSL (http://www.fmrib.ox.ac.uk/fsl/) was used in the cross-sectional pipeline. The deformation matrix, which was automatically calculated within the longitudinal stream, was applied to both the test and the retest maps.

2.3.2. *Nuclei of the amygdala and hippocampal subfields segmentation*

The module for nuclei of the amygdala and hippocampal subfields, available with the developmental version freesurfer-x86_64-unknown-linux-gnu-dev-20170915 (https://surfer.nmr.mgh.harvard.edu/fswiki/HippocampalSubfieldsAndNucleiOfAmygdala), was applied to images processed with both the cross-sectional and the longitudinal pipelines. The algorithm uses a Bayesian inference, where each voxel is labeled combing the probabilistic atlas and image intensities (Iglesias et al., 2015; Saygin et al., 2017). For both amygdalar nuclei and hippocampal subfields, the probabilistic atlas was built combining *ex vivo* manually labeled (nuclei of the amygdala/hippocampal subfields) and *in vivo* manually segmented (outer contours and surrounding structures) data (Iglesias et al., 2015; Saygin et al., 2017). The atlases are mapped to the subject image using an affine, robust registration (Reuter et al., 2010). All voxels not covered by the mapped atlas are not considered for the segmentation. For the cross-sectional processing, the skull-stripped and intensity corrected volume produced as output from the standard pipeline is used to initialize the nuclei and subfields segmentation (Iglesias et al., 2015; Saygin et al., 2017).

Similarly to the whole brain segmentation, the longitudinal pipeline for the amygdalar nuclei and hippocampal subfields (Iglesias et al., 2016) uses as input the normalized, intensity corrected, skull-stripped images from the longitudinal standard processing. The binary masks of the amygdala and the hippocampus, extracted from the subject template using a soft segmentation, are used as reference for the deformation to the atlas.

All the cortical and subcortical segmentations (both from the cross-sectional and the longitudinal pipeline) were visually inspected, but no manual edit was carried out to avoid the introduction of any external bias. Two subjects, one from the older cohort and one from the younger cohort, were excluded from further analysis due to a segmentation failure. Moreover, 4 subjects were excluded from the T1-averaged sample (longitudinal pipeline) due to the failure to segment the nuclei of the amygdala and the hippocampal subfields in the right hemisphere. Within each cohort, and for each structure and time-point, the presence of extreme outliers (defined as values 1.5 fold above the interquartile range)

was assessed (IBM SPSS Statistics for Windows, version 23), and the related measures were excluded from the statistical analysis. In the older cohort, the final sample size was $N = 64$ for the sample with the MRI acquisition at the beginning of the session ("scan 1" sample), $N = 60$ for the sample with the MRI acquisition at the end of the session ("scan 2" sample), and $N = 56$ for the T1-averaged sample. Sixty-seven subjects were included in the younger cohort.

2.4. Reliability analysis

The reliability analysis was performed by computing the percent absolute reproducibility error (ε, in percent) and the spatial reproducibility coefficient (DICE) across the test-retest session (Jovicich et al., 2013; Marizzoni et al., 2015), which were calculated as follows:

$$\varepsilon = 100 \ \mathrm{x} \frac{|V_{retest} - V_{test}|}{(V_{retest} + V_{test})/2},$$

$$DICE = 2 \frac{|M_{retest}| \ |M_{test}|}{|M_{retest}| + |M_{test}|}$$

where V_{retest} and V_{test} are the volumes of nuclei/subfields and M_{retest} and M_{test} are the binary 3D masks of nuclei/subfields from the two different MRI sessions. While ε provides information about the variability of volumetric measures (percentage difference of volumes estimation), the DICE is an overlap index evaluating the spatial reproducibility of 3D volumes (i.e., the spatial accuracy) (Taha and Hanbury, 2015).

The Interclass Correlation Coefficient (ICC) (Shrout and Fleiss, 1979) was included as further index of absolute agreement between the test–retest measures ($ICC_{2,1}$) (Rajaratnam, 1960). The ICC calculated (IBM SPSS Statistics for Windows, version 23) for all volumes of both the cohorts.

All analyses used raw volumes instead of volumes corrected for total intracranial volume.

2.5. Statistical analysis

Non-parametric statistic tests were used since they are more appropriate for small samples, for which normality of data distributions cannot be assumed. Statistical analyses were conducted using IBM SPSS Statistics for Windows, version 23. One-way Kruskal–Wallis tests (nonparametric version of ANOVA) were used to: i) compare the demographic features between sites, and ii) to test for potential effects of MRI site and vendor on reproducibility measures. Wilcoxon signed rank tests were performed to: i) compare the reliability of the cross-sectional *vs*. the longitudinal pipelines, and ii) test the effect of averaging two T1-weighted images (older cohort) compared to the use of a single scan (scan 1 and scan 2) on reproducibility metrics. The Mann-Whitney *U* test was used when comparisons involved only two groups. Finally, the Spearman's rank correlation was computed to test the association between volumes and reproducibility metrics.

The reliability analysis was performed considering each nucleus and subfield as a single structure (anatomical parcellation, Supplementary Table 1), as well as grouping them into their relevant subregions (nuclei and subfields' groups, Supplementary Table 1). In particular, as regards for the amygdalar nuclei, to overcome the variability across grouping schemes, we included the basal, the lateral, and accessory basal nuclei within the basolateral complex, and the central and the medial nuclei in the centromedial complex, as these were the regions with the greatest agreement across studies. Moreover, we tested whether the test-retest reproducibility was affected by averaging two structural T1 scans from the same session.

Finally, we provided estimates of the sample size required for a hypothetical longitudinal study on hippocampal atrophy in Alzheimer's disease (AD) patients, assuming a 1:1 enrollment ratio between patients and controls, a 0.85 power to reject the null hypothesis that the mean volumes of the AD and controls are equal, and a 0.05 Type I error probability (α).The sample size was calculated using PS Power and Sample Size Calculations tool, v 3.0 (Dupont and Plummer Jr, 1990; Dupont and Plummer Jr, 1998; http://biostat.mc.vanderbilt.edu/PowerSampleSize).

3. Results

Fig. 1 shows the automatic segmentation derived from a representative older (A) and younger (B) volunteer, respectively. In the older cohort (age range: 50–78 years), subjects from Site 2 were older compared to those acquired at Sites 5, 6, 8 (for all, Mann-Whitney test: $U = 0, p = .01$), 9 ($U = 2$, $p = .03$), and 12 ($U = 3$, $p = .05$), and Site 5 volunteers were younger compared to those from Sites 2, 7 (for all, $U = 0, p = .01$), and 11

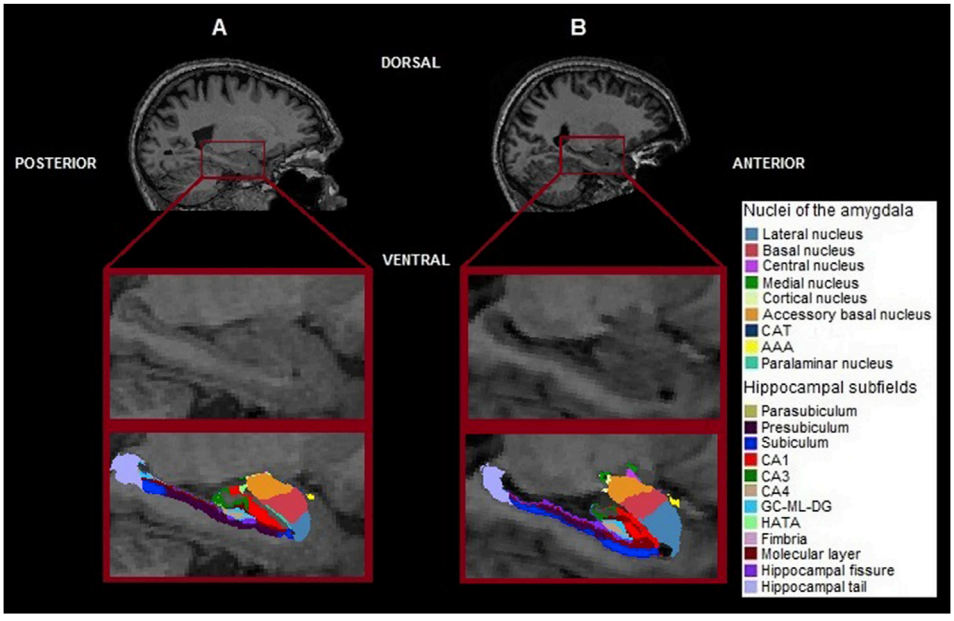
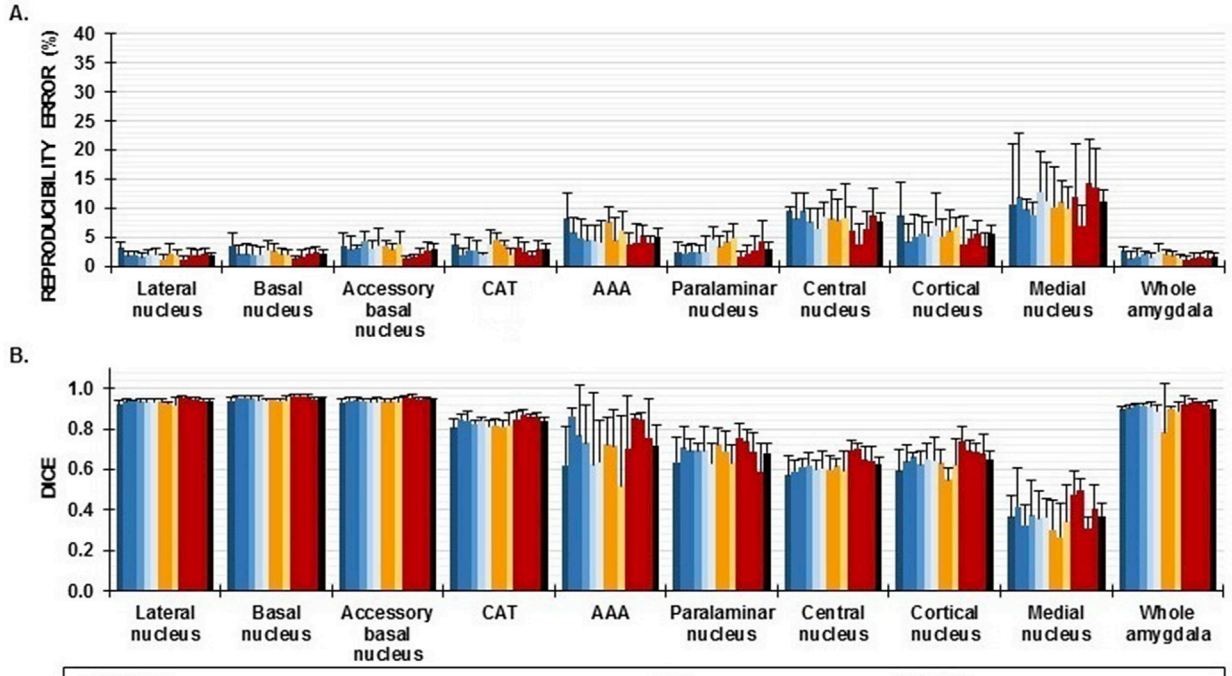


Fig. 1. Nuclei of the amygdala and hippocampal subfields segmentations output from FreeSurfer v6.0 (left hemisphere) from one representative older (A) and younger (B) subject.

Table 2
Mean values of volume reproducibility error (ε), DICE, and ICC coefficients for the nuclei of the amygdala and hippocampal subfields in the older and the younger cohorts. Values are averaged across hemispheres, structures and subjects.

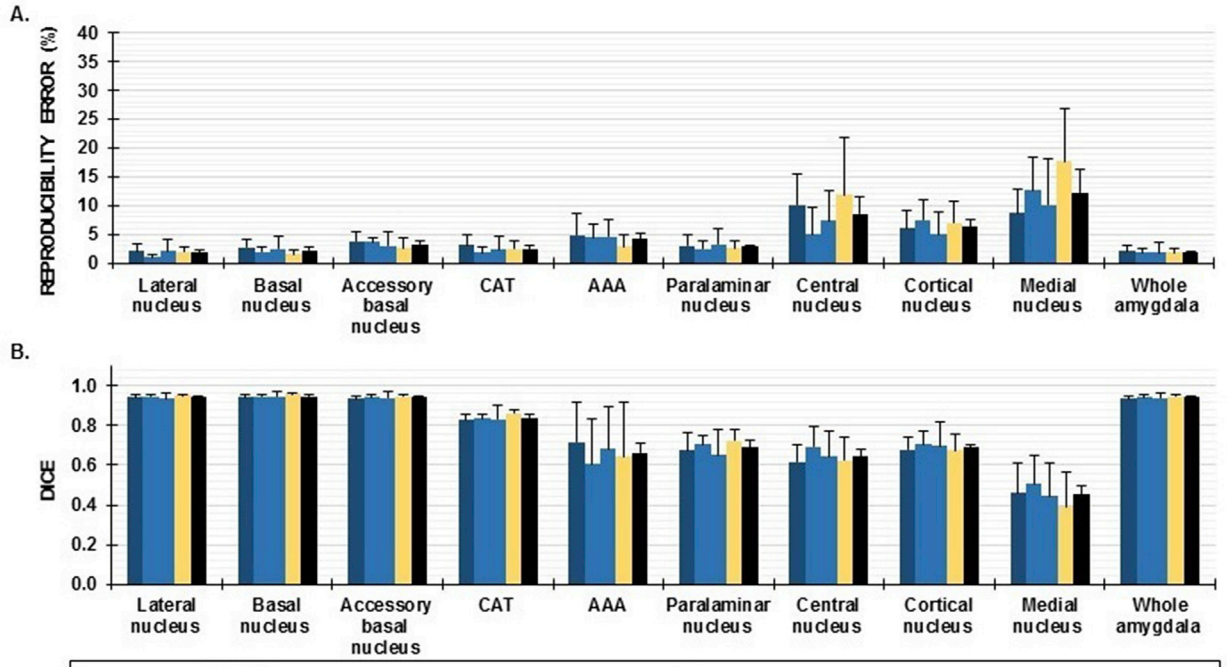
Amygdala	M(SD)	ε[a] Min–max (MRI scanner)	DICE[b]		ICC[b]	
			M(SD)	Min–max (MRI scanner)	M(SD)	Min–max (MRI scanner)
Anatomical parcellation - all nuclei						
Older (scan 1)	4.48% (0.63)	3.15% (Philips Achieva) −5.66% (Siemens Allegra)	0.76 (0.03)	0.72 (GE Discovery) −0.82 (Philips Achieva)	0.92 (0.05)	0.82 (Siemens TrioTim) – 0.98 (Siemens Skyra)
Younger	4.61% (0.48)	4.23% (Siemens TrioTim) −5.26% (GE Discovery)	0.77 (0.01)	0.77 (Siemens TrioTim and Allegra, GE Discovery) −0.78 (Siemens TrioTim)	0.96 (0.02)	0.93 (GE Discovery) – 0.97 (Siemens Allegra and TrioTim)
Hippocampus						
Anatomical parcellation – all subfields						
Older (scan 1)	3.46% (0.74)	2.58% (Philips Achieva) −4.74% (Siemens Allegra)	0.82 (0.03)	0.78 (Siemens Allegra) −0.86 (Philips Achieva)	0.96 (0.03)	0.90 (GE HDxt) – 0.98 (Philips Achieva)
Younger	3.21% (0.34)	2.79% (Siemens TrioTim) −3.52% (Siemens Allegra)	0.83 (0.02)	0.81 (Siemens Allegra) −0.85 (GE Discovery)	0.97 (0.01)	0.96 (Siemens TrioTim) – 0.98 (Siemens TrioTim, GE Discovery)

Abbreviations: M, mean; SD, standard deviation; min, minimum value; max, maximum value.
[a] Range 0–100, with 0 as the best score.
[b] Range 0–1, with 1 as the best score.



	MRI site effect				MRI vendor effect			
	ε		DICE		ε		DICE	
Nuclei of the amygdala	*H*(*df*,12)	*p*	*H*(*df*,12)	*p*	*H*(*df*,2)	*p*	*H*(*df*,2)	*p*
Lateral nucleus	12.52	.40	25.11	.01	1.75	.42	18.15	.0001
Basal nucleus	9.88	.63	22.78	.03	1.84	.40	12.68	.002
Accessory basal nucleus	17.08	.15	19.90	.07	9.77	.01	16.86	.0002
CAT	16.61	.17	21.12	.05	4.67	.10	13.41	.001
AAA	12.13	.44	24.33	.02	3.90	.14	8.35	.02
Paralaminar nucleus	14.99	.24	12.26	.43	8.08	.02	1.60	.45
Central nucleus	10.44	.58	19.27	.08	4.16	.13	13.59	.001
Cortical nucleus	9.00	.70	16.34	.18	1.69	.43	9.47	.01
Medial nucleus	5.43	.94	14.40	.28	0.14	.93	3.63	.16
Whole amygdala	17.12	.15	30.91	.002	4.05	.13	20.92	.00003

Fig. 2. Across session volume reproducibility error (ε) (A) and DICE coefficient (B) of whole and nuclei of amygdala in the older volunteers. Error bars denote the standard deviation of the mean, computed averaging the test-retest ε and DICE across subjects and hemispheres. The black column denotes the averaged values across MRI site. *H* and *p* in the Table denote the chi-square and *p* value for the Kruskal-Wallis test, respectively.
Abbreviations: *df*, degrees of freedom.



	MRI site effect				MRI vendor effect			
	ε		DICE		ε		DICE	
Nuclei of the amygdala	*H*(*df*,3)	*p*	*H*(*df*,3)	*p*	*H*(*df*,1)	*p*	*H*(*df*,1)	*p*
Lateral nucleus	4.38	.22	3.14	.37	0.25	.62	2.80	.09
Basal nucleus	6.06	.11	8.97	.03	2.99	.08	3.45	.06
Accessory basal nucleus	6.94	.07	5.20	.16	0.95	.33	1.06	.30
CAT	5.55	.14	6.50	.09	0.07	.80	4.32	.04
AAA	2.67	.45	2.05	.56	2.62	.11	0.57	.45
Paralaminar nucleus	0.35	.95	3.65	.30	0.16	.69	2.92	.09
Central nucleus	7.03	.07	4.45	.22	0.39	.53	0.06	.81
Cortical nucleus	6.54	.09	5.21	.16	1.52	.22	1.10	.30
Medial nucleus	9.24	.03	1.83	.61	7.31	.01	1.18	.28
Whole amygdala	3.08)	.38	3.58	.31	0.06	.81	1.30	.25

Fig. 3. Across sessions volume reproducibility error (ε) (A) and DICE coefficient (B) of whole and nuclei of amygdala in the younger volunteers. Error bars denote the standard deviation of the mean, computed averaging the test-retest ε and DICE across subjects and hemispheres. The black column denotes the averaged values across MRI sites. *H* and *p* in the Table denote the chi-square and *p* value for the Kruskal-Wallis test, respectively. Abbreviations: *df*, degrees of freedom.

($U = 2$, $p = .03$) (Table 1). In the younger cohort (age range: 18–43 years), subjects from Site 16 were younger than those from Site 14 ($U = 81$, $p = .000003$), 15 ($U = 29$, $p = .02$) and 17 ($U = 49.5$, $p = .002$) (Table 1). No differences in sex distribution were observed in either cohort. The test-retest interval was of 17 ± 12 days (range 7–56) in the older group and 31 ± 26 days (range 1–89) in the younger group.

3.1. Cross-sectional vs longitudinal pipeline reproducibility

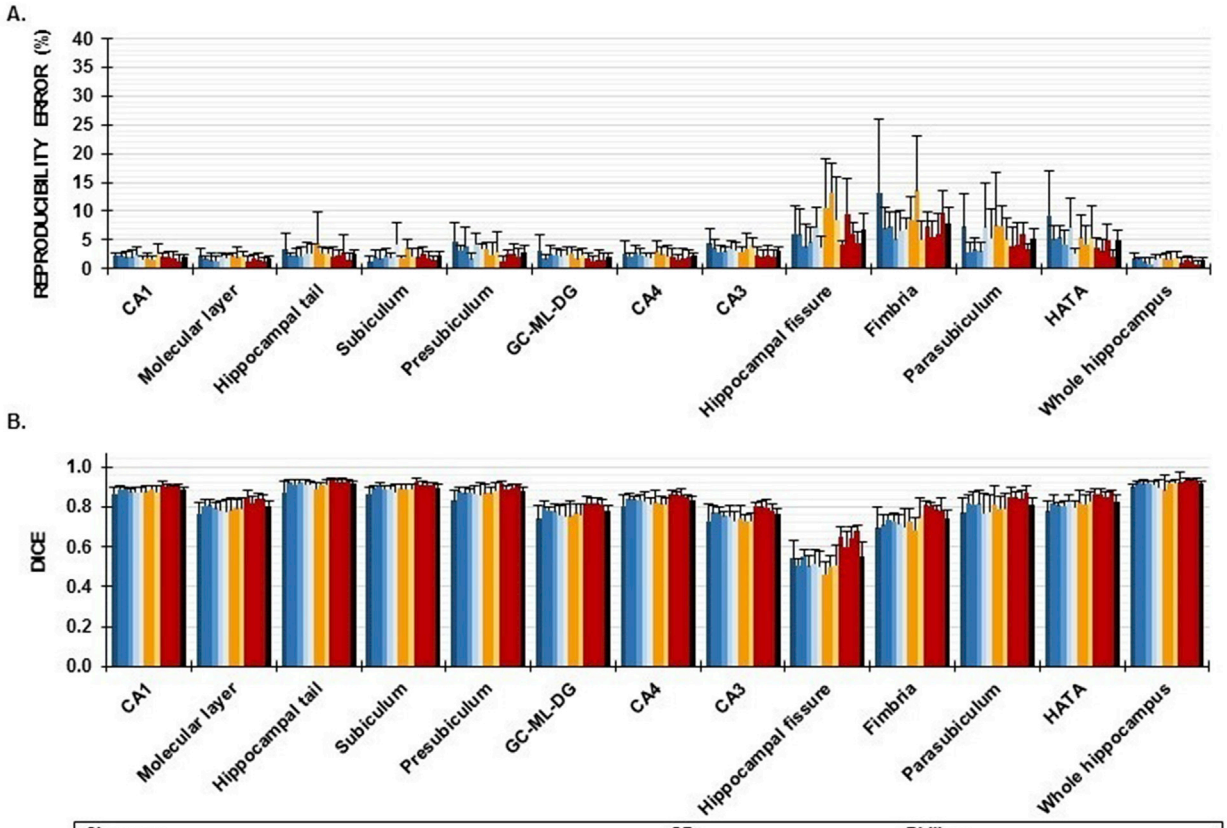
The longitudinal pipeline showed lower ε and higher DICE and ICC values compared to the cross-sectional pipeline in the older cohort, regardless of the within-session scan used, and, in absolute terms, in the younger cohort (Supplementary Fig. 1). All the results reported below refer to the segmentations derived from the longitudinal pipeline and, for the older cohort, to the scan acquired at the beginning of the MRI session (scan 1) (scan 1 *vs* scan 2, longitudinal pipeline: Wilcoxon test: ε, $Z = 0.31$, $p = .75$; DICE, $Z = 0.80$, $p = .42$; ICC, $Z = 0.25$, $p = .81$).

3.2. Reproducibility of the longitudinal pipeline: effect of site, vendors and nuclei/subfields grouping

Reproducibility metrics are reported in Table 2 (anatomical parcellation) and in the Supplementary Table 2 (nuclei and subfields' groups). Focusing on the nuclei of the amygdala, the larger nuclei (i.e. from the lateral nucleus to the CAT) showed a mean ε lower than 5% as well as a mean DICE higher than 0.80 (Fig. 2, Fig. 3). The effect of MRI sites and vendors were reported mainly for the DICE of the older cohort (Figs. 2 and 3). Focusing on hippocampal subfields, the larger subfields (i.e. from the CA1 to the presubiculum, except for the molecular layer) showed a mean ε lower than 5% as well as a mean DICE higher than 0.80 (Fig. 4, Fig. 5). The effect of MRI sites and vendors were extensively observed for the DICE in both cohorts, and to a limited extent for the ε (Figs. 4 and 5). The effect of different MRI sites/vendors on DICE was limited by the nuclei and subfields' grouping only in the younger cohort (Supplementary Fig. 2). Along the same line, larger volumes of the nuclei of the amygdala and of the hippocampal subfields were associated with lower ε as well as higher DICE as revealed by Spearman's correlation (Fig. 6). Similar results were found when nuclei and subfields were grouped in the respective relevant components (Supplementary Fig. 3).

3.3. Effect of the within-session T1 averaging

Within-session T1 averaging significantly improved the volumetric reliability of hippocampal subfields segmentations and, to a limited extent,



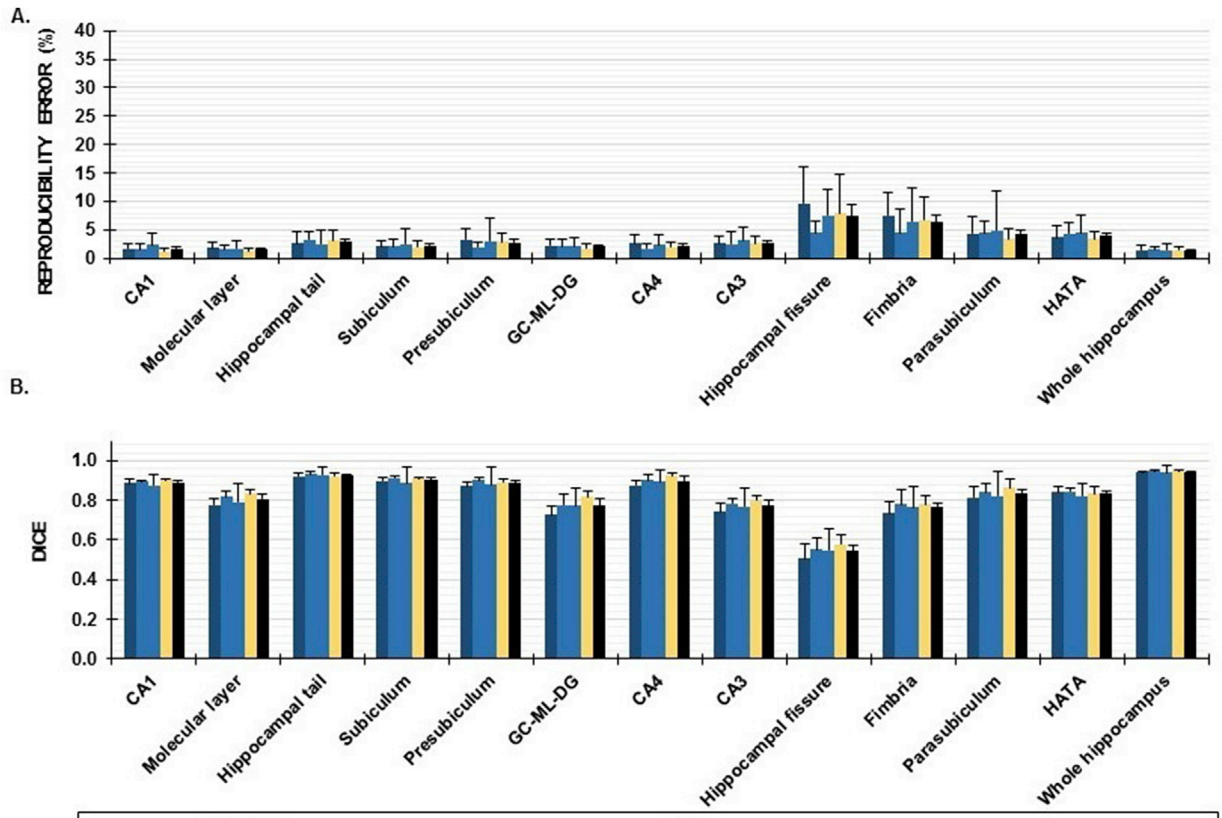
	MRI site effect				MRI vendor effect			
	ε		DICE		ε		DICE	
Hippocampal subfields	*H*(*df*,12)	*p*	*H*(*df*,12)	*p*	*H*(*df*,2)	*p*	*H*(*df*,2)	*p*
CA 1	10.02	.61	28.48	.01	0.85	.65	22.91	.00001
Molecular layer	11.61	.48	21.25	.05	3.95	.14	16.72	.0002
Hippocampal tail	4.59	.97	27.11	.01	2.57	.28	17.50	.0002
Subiculum	16.77	.16	29.33	.004	1.90	.39	22.14	.00002
Presubiculum	18.12	.11	22.61	.03	5.23	.07	16.21	.0003
GC-ML.DG	9.63	.65	20.88	.05	6.74	.03	17.12	.0002
CA 4	6.29	.90	16.71	.16	4.25	.12	12.93	.002
CA 3	13.13	.36	23.58	.02	7.56	.02	20.38	.00004
Hippocampal fissure	17.57	.13	39.26	.0001	5.82	.05	29.60	3.75E-7
Fimbria	11.59	.48	30.02	.003	1.36	.51	28.48	6.55E-7
Parasubiculum	9.04	.70	16.48	.17	2.71	.26	11.02	.004
HATA	13.37	.34	30.02	.003	2.83	.24	23.53	.000008
Whole hippocampus	14.90	.25	27.99	.01	2.82	.24	23.09	.00001

Fig. 4. Across sessions volume reproducibility error (ε) (A) and DICE coefficient (B) of whole hippocampus and hippocampal subfields in the older volunteer. Error bars denote the standard deviation of the mean, computed averaging the test-retest ε and DICE across subjects and hemispheres. The black column denotes the averaged values across MRI sites. *H* and *p* in the Table denote the chi-square and *p* value for the Kruskal-Wallis test, respectively.
Abbreviations: *df*, degrees of freedom.

of the amygdalar nuclei. The effect emerged also for the spatial reproducibility of both nuclei of the amygdala and hippocampal subfields segmentations, except for the DICE of the hippocampal head in the subfields grouping, where decreased reliability after the T1 averaging was found (Supplementary Table 2, Supplementary Fig. 4; Supplementary Fig. 5; Supplementary Fig. 7). Similar to ε, the ICC was affected in a limited manner by the within-session T1-averaging (Supplementary Fig. 6). Despite the increase of the reproducibility indices, the MRI site effect was still evident for the DICE, while the MRI vendor effect was still detected for both the ε and the DICE (Supplementary Fig. 4; Supplementary Fig. 5), but impacting a lower number of structures. Similar results emerged when the reliability analysis was performed on the nuclei and subfields' groups (Supplementary Fig. 6; Supplementary Fig. 7).

3.4. Sample size estimation

We focused on AD patients as hippocampal atrophy is one of the hallmarks of the disease and longitudinal rates of atrophy are well documented. Here, we hypothesised an annual (i.e., over 12 months) atrophy rate of 4.51% ± 3.06 for the whole hippocampus (Barnes et al., 2009). Moreover, we considered the CA1 that is the most affected subfield in AD (Frankó et al., 2013). Annual atrophy rates are not available for the CA1, thus we hypothesised a range of values assuming that CA1 atrophy rates would be more pronounced than the whole hippocampus atrophy rates. In particular, we considered the following atrophy rates: 4.51% + 0.5 standard deviation (SD) (i.e., 6.04%), 4.51% + 1 SD (7.60%), 4.51% + 1.5 SD (9.10%), 4.51% + 2 SD (10.60%). In the older cohort, the mean volumes were 3353 ± 244 mm^3 and 648 ± 54 mm^3 for



Siemens: ■ Allegra	■ TrioTim (2)	GE: ■ Discovery	■ Mean across sites

	MRI site effect				MRI vendor effect			
	ε		DICE		ε		DICE	
Hippocampal subfields	*H*(*df*,3)	*p*	*H*(*df*,3)	*p*	*H*(*df*,1)	*p*	*H*(*df*,1)	*p*
CA 1	5.01	.17	3.92	.27	3.87	.05	3.45	.06
Molecular layer	2.22	.53	14.03	.003	1.18	.28	5.63	.02
Hippocampal tail	3.63	.31	11.12	.01	1.10	.30	2.50	.11
Subiculum	0.38	.94	4.85	.18	0.16	.69	0.28	.59
Presubiculum	7.79	.05	13.56	.004	1.26	.26	0.00	.97
GC-ML.DG	0.82	.85	18.50	.0003	0.60	.44	7.41	.01
CA 4	1.90	.59	18.36	.0004	0.27	.61	7.91	.01
CA 3	0.83	.84	9.25	.03	0.02	.88	2.27	.13
Hippocampal fissure	3.74	.29	8.91	.03	0.09	.77	1.85	.17
Fimbria	4.55	.21	10.24	.02	0.21	.65	0.14	.71
Parasubiculum	1.87	.60	8.22	.04	0.18	.67	2.92	.09
HATA	0.91	.82	1.33	.72	0.20	.66	0.08	.78
Whole hippocampus	1.84	.61	9.93	.02	0.15	.70	0.18	.67

Fig. 5. Across sessions volume reproducibility error (ε) (A) and DICE coefficient (B) of whole hippocampus and hippocampal subfields in the younger volunteers. Error bars denote the standard deviation of the mean, computed averaging the test-retest reproducibility error and DICE across subjects and hemispheres. The black column denote the averaged values across MRI sites. *H* and *p* in the Table denote the chi-square and *p* value for the Kruskal-Wallis test, respectively.
Abbreviations: *df*, degrees of freedom.

the whole hippocampus and the CA1, respectively. As expected, the sample size decreased with longer follow-up periods, and the advantage of using the CA1 emerged for atrophy rates as high as 9.10% (4.51% + 1.5 SD) (Supplementary Fig. 8).

4. Discussion

In the present study, we evaluated the test-retest reliability of the FreeSurfer-based automated segmentation of hippocampal subfields and, for the first time, of amygdalar nuclei in two independent cohorts of healthy volunteers. The reliability was assessed considering the single nuclei/subfields (anatomical parcellation) and grouping them by relevant components. Our results showed that: i) differences in MRI sites/vendors had a limited effect on volume reproducibility; ii) differences in MRI sites/vendors had non-negligible effect on spatial reproducibility; iii) the reliability was strongly influenced by the absolute volume size; and iv) the within-session T1-averaging improved both the volume and the spatial reproducibility, also partially limiting the effect of different MRI sites/vendors, as it affected a lower number of structures.

In the older cohort, we found a significantly lower ε, higher DICE, and higher ICC for the longitudinal as compared to the cross-sectional pipeline, while the difference was not statistically significant for the younger cohort. Results from the older cohort are in line with previous reports on the same topic (Marizzoni et al., 2015; Worker et al., 2018). For the younger cohort, results are similar to those from the recent work from Brown and colleagues (Brown et al., 2020), which reported reliable

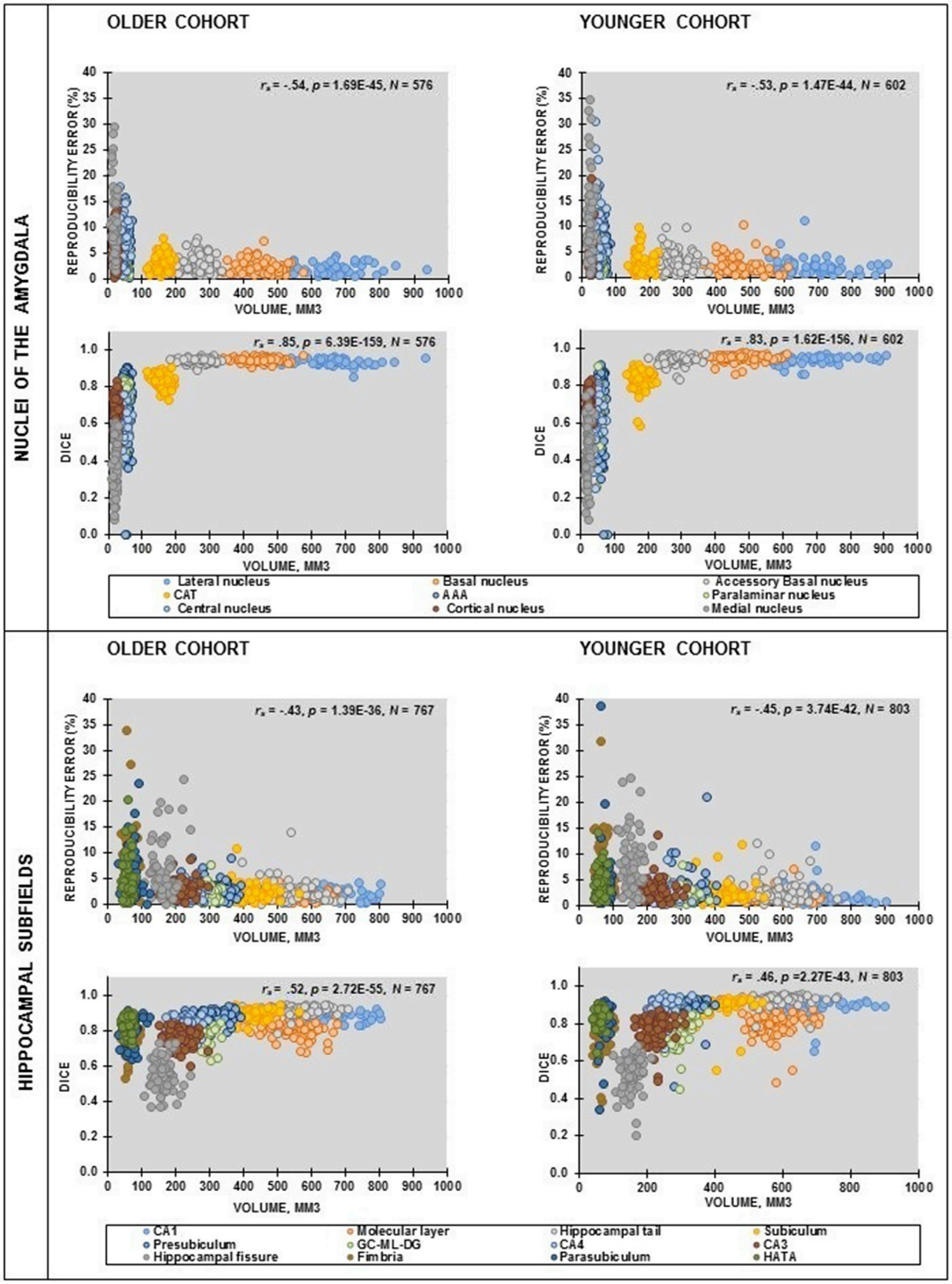


Fig. 6. Relationship between the volumes of the nuclei of the amygdala and hippocampal subfields and the reproducibility error and the DICE coefficient in the older and younger cohorts. Each dot denotes a subject. Within subjects' volumes were averaged across hemisphere and sessions.

hippocampal subfields estimates over 3 time-points; however, we note we lack the terms for comparison for the amygdala, since no study investigated the effect of the pipeline on the reliability of these structures in this population. The effect of the longitudinal pipeline is probably related to the co-registration of all subject's images to a common space (Reuter et al., 2012).

In line with previous reports on hippocampal subfields (Marizzoni et al., 2015; Yushkevich et al., 2015), we found a good volume reliability (i.e., mean ε lower than 5%) and spatial reproducibility (i.e., mean DICE $\geq$ 0.76, where a DICE $>$ 0.70 is an index of good spatial overlap; Zijdenbos et al., 1994; Zou et al., 2004). The ICC $>$0.92 denotes an excellent consistency between segmentations, where an ICC $\geq$0.75 is considered an index of strong agreement (Cicchetti, 1994). Our results are also consistent with previous works investigating hippocampal subfields segmentation using tools different from FreeSurfer. For example, a recent work by Amaral and colleagues (Amaral et al., 2018) performed an extensive validation of a fully automated pipeline (MAGeT Brain tool; Pipitone et al., 2014) for the segmentation of whole hippocampus and subfields (CA1, CA2/CA3, dentate gyrus/CA4, subiculum) and related white matter projections (i.e., alveus, fimbria, and fornix), including ICC and DICE as reliability measures. Considering only the hippocampal subfields that have been investigated by both our and Amaral's studies, we generally reported greater or equal values of DICE and mean ICC.

Similar to the main analysis, a good volume and spatial reproducibility and across measures agreement emerged when considering the reliability of the nuclei and subfields' groups, except for the ε of the amygdala, probably due to the inclusion of the smaller nuclei (central and medial nuclei, for the centromedial complex). However, this is consistent with the assumption of an inverse relationship between volume size and ε, which was also confirmed in this study. In general, considering a ε lower than 5% and a DICE $\geq$0.80 as indicative of good reproducibility, these data suggest that nuclei larger than 200 mm^3 and subfields larger than 300 mm^3 were the most reliable, both in terms of volume and spatial measures.

Our results revealed that differences in MRI sites and vendors had a limited impact on volumetric reproducibility measure ε, while they strongly influenced the spatial reliability (DICE). In the younger cohort, the effect of different MRI sites/vendors on DICE was milder when considering the nuclei and subfields' groups. One possible explanation of the different impact of the MRI sites/vendors on the volume and the spatial reliability could be related to the anatomical complexity of these structures. In particular, the anatomical borders between the amygdala and the hippocampus do not reliably correspond to macroscopically visible landmarks, which constitute the basis for MRI-based definition of these structures (Amunts et al., 2005; Konrad et al., 2009). Taken together, these results support the usage of volumetric measures in multicentric MRI studies, while those spatial should be used more carefully, particularly in the old population. As reported in previous studies (Marizzoni et al., 2015; Morey et al., 2010; Van Leemput et al., 2009; Worker et al., 2018), the segmentation reliability was strongly influenced by the absolute volumes of the respective structures. Indeed, we observed a strong negative correlation between volume and ε and a strong positive association between volume and DICE, irrespective of cohort and region. Similar results were reported when considering the nuclei and subfields' groups, suggesting that data derived from smaller nuclei/subfields should be interpreted very cautiously.

Finally, in the older cohort we evaluated the effect of the within-session T1 averaging on the test-retest reproducibility of the amygdalar nuclei and the hippocampal subfields. The effect of the within-session T1 averaging had previously been assessed in samples of healthy old volunteers from the PharmaCog Consortium (Jovicich et al., 2013; Marizzoni et al., 2015), and the advantage of T1 averaging was evident for the reliability of hippocampal subfields (Marizzoni et al., 2015). Our results confirm the higher reproducibility of measures obtained on averaged T1 images for hippocampal subfields, and expand this observation to amygdalar nuclei and to nuclei/subfields' groups. The superiority of T1 averaging was mainly detected for the DICE of almost all structures. The MRI site/vendor effect was detected also for T1-averaged images, but involved a lower number of structures. Thus, the within-session T1 averaging could be a valuable strategy to partially limit the MRI sites/vendors effect in multicentric studies. In this context, however, an important issue should be considered, i.e. the quality of both images needs to be good otherwise the averaging would not improve the segmentation reliability.

The amygdala and the hippocampus are known to play a key role in neurodegenerative disease (e.g., AD) and several psychiatric conditions, such as schizophrenia (Ganzola et al., 2014; Prestia et al., 2015; Rich et al., 2016; Bartsch et al., 2019), bipolar disorder (Janiri et al., 2017; Wijeratne et al., 2013), major depressive disorder (Cullen et al., 2014; Leal et al., 2017), as well as borderline personality disorder (Driessen et al., 2000; Schmahl et al., 2009; Weniger et al., 2009), and several studies supported the involvement of specific nuclei and subfields (Asami et al., 2018; Cao et al., 2017; Gryglewski et al., 2019). Within this framework, our results would be useful for future multicentric neuroimaging studies in the field of psychiatry and dementia. In particular, the possibility to obtain reliable segmentation of nuclei/subfields with the advantage in terms of processing time compared to manual segmentation could facilitate the study of structural abnormalities and treatment responses in large multicentric dataset. Indeed, these nuclei/subfields could be used as diagnostic markers or as outcome markers in future clinical trials of potentially disease-modifying therapies. This work has several limitations. The major limitation was that, in the older cohort, the number of enrolled volunteers was only 5 per site, as already discussed in other articles on the PharmaCog project (Jovicich et al., 2013, 2014; Marizzoni et al., 2015). Comparisons across MRI sites allowed us to assess the effect of the individual variability, which is a relevant issue in multicentric studies. To overcome the limited sample size per site and to obtain results which were generalizable to manufacturers, we also tested the segmentations' reliability grouping subjects by MRI vendors. The second limitation was that the set of volunteers scanned at each MRI site was different. However, this homogeneity should not affect our results as each subject acted as his/her own control in the test-retest session, and the aim of this study was to evaluate the reliability of the segmentation and not structural differences between groups. Finally, the MRI protocol was designed for the clinical setting and it did not include a high-resolution T1 acquisition neither a 3D T2 sequence. The T2 sequence has become very popular in neuroimaging studies, as T2 acquisitions in the coronal plane, perpendicular to the main axis of the hippocampus, facilitate the identification of the molecular layer and of its boundary with the CA1, which are usually not clearly visible on standard T1 data (Iglesias et al., 2015; Wisse et al., 2014). Moreover, the multimodal MRI approach, such as combining 3D T1 and 3D FLAIR sequences, was reported to improve segmentation of the hippocampus (Viviani et al., 2017).

Notwithstanding the advantages of T2 sequences, our results could still be useful for clinical and research centers using a standard MRI protocol, while future studies are needed to assess the reliability of these measures using more advanced protocols. A recent study (Mueller et al., 2018) compared the hippocampal subfields segmentation of four automated protocols (both T1-based and T2-based) and the manual technique. The study showed that high-resolution T2 approaches performed better than T1 approaches in the hippocampal subfields segmentation. Further studies should assess whether this difference also applies to the nuclei of the amygdala. A limitation of the younger cohort was the lack of within-session T1 acquisitions, which precluded the possibility to assess the effect of averaging on reproducibility within this group. Though, the assessment of the effect of within-session T1 averaging on reliability measures was not a primary aim of this study. Albeit preliminary, our result of a potential benefit of within-session T1 averaging on FreeSurfer-based segmentation warrants further investigation. Finally, the MRI acquisition protocols were not harmonized for the younger cohort. In particular, a possible question is whether different voxel-sizes

(1.3x1x1.3 mm *vs* 1.3x1x1mm *vs* 1.3x1x1mm *vs* 1 mm isotropic) could affect the reliability of segmentations. However, an MRI site effect was also found in the older cohort, where the MRI protocols have been previously harmonized (Jovicich et al., 2013) and voxel-size was the same between MRI centers.

Conclusions

Our results support the use of the FreeSurfer v6.0 for amygdalar nuclei and hippocampal subfields segmentations from MRI brain images in the context of multicenter studies. Specifically, our study suggests that larger nuclei of the amygdala and hippocampus are reliable and could be used to assess disease progression and treatment response in future research trials. In particular, these measures could be valuable in volumetric analysis due to the limited effect of site on volume reproducibility, as assessed by the analysis of the reproducibility error. On the other hand, the usage of these measures should be carefully considered in the context of spatial-based analyses, due to the higher sensibility of the spatial reproducibility to which the DICE is referred) to MRI scan features. To limit this effect, possible alternatives are grouping the amygdalar nuclei and the hippocampal subfields, and using the within-session T1-averaged images.

Funding

PharmaCog is funded by the EU-FP7 for the InnovativeMedicine Initiative (grant no. 115009).

Declaration of competing interest

None.

CRediT authorship contribution statement

Giulia Quattrini: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Michela Pievani:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Jorge Jovicich:** Investigation, Writing - review & editing. **Marco Aiello:** Investigation, Writing - review & editing. **Núria Bargalló:** Investigation, Writing - review & editing. **Frederik Barkhof:** Investigation, Writing - review & editing. **David Bartres-Faz:** Funding acquisition, Resources, Writing - review & editing. **Alberto Beltramello:** Investigation, Writing - review & editing. **Francesca B. Pizzini:** Investigation, Writing - review & editing. **Olivier Blin:** Funding acquisition, Resources, Writing - review & editing. **Regis Bordet:** Project administration. **Massimo Caulo:** Investigation, Writing - review & editing. **Manos Constantinides:** Investigation, Writing - review & editing. **Mira Didic:** Funding acquisition, Resources, Writing - review & editing. **Antonios Drevelegas:** Investigation, Writing - review & editing. **Antonio Ferretti:** Investigation, Writing - review & editing. **Ute Fiedler:** Investigation, Writing - review & editing. **Piero Floridi:** Investigation, Writing - review & editing. **Hélène Gros-Dagnac:** Investigation, Writing - review & editing. **Tilman Hensch:** Investigation, Writing - review & editing. **Karl-Titus Hoffmann:** Investigation, Writing - review & editing. **Joost P. Kuijer:** Investigation, Writing - review & editing. **Renaud Lopes:** Investigation, Writing - review & editing. **Camillo Marra:** Funding acquisition, Resources, Writing - review & editing. **Bernhard W. Müller:** Investigation, Writing - review & editing. **Flavio Nobili:** Funding acquisition, Resources, Writing - review & editing. **Lucilla Parnetti:** Funding acquisition, Resources, Writing - review & editing. **Pierre Payoux:** Funding acquisition, Resources, Writing - review & editing. **Agnese Picco:** Investigation, Writing - review & editing. **Jean-Philippe Ranjeva:** Investigation, Writing - review & editing. **Luca Roccatagliata:** Investigation, Writing - review & editing. **Paolo M. Rossini:** Funding acquisition, Resources, Writing - review & editing. **Marco Salvatore:** Investigation, Writing - review & editing. **Peter Schonknecht:** Funding acquisition, Resources, Writing - review & editing. **Björn H. Schott:** Investigation, Writing - review & editing. **Julien Sein:** Investigation, Writing - review & editing. **Andrea Soricelli:** Funding acquisition, Resources, Writing - review & editing. **Roberto Tarducci:** Investigation, Writing - review & editing. **Magda Tsolaki:** Funding acquisition, Resources, Writing - review & editing. **Pieter J. Visser:** Funding acquisition, Resources, Writing - review & editing. **Jens Wiltfang:** Funding acquisition, Resources, Writing - review & editing. **Jill C. Richardson:** Project administration. **Giovanni B. Frisoni:** Funding acquisition, Resources, Writing - review & editing, Supervision. **Moira Marizzoni:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing, Supervision.

Acknowledgements

PharmaCog is funded by the EU-FP7 for the Innovative Medicine Initiative (grant no. 115009). We wish to thank all the co-Investigators from our Study Group (WP5-PharmaCog/European Alzheimer's Disease Neuroimaging Initiative) (for the complete list see Supplementary Table 3). Moreover, we are grateful to all members and collaborators of the PharmaCog project, and particularly to Luigi Antelmi and Daniele Orlandi.

FB is supported by the NIHlgR biomedical research center at UCLH.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116932.

References

Amaral, R.S., Park, M.T.M., Devenyi, G.A., Lynn, V., Pipitone, J., Winterburn, J., Chavez, S., Schira, M., Lobaugh, N.J., Voineskos, A.N., 2018. Manual segmentation of the fornix, fimbria, and alveus on high-resolution 3T MRI: application via fully-automated mapping of the human memory circuit white and grey matter in healthy and pathological aging. Neuroimage 170, 132–150.

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N., Habel, U., Schneider, F., Zilles, K., 2005. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. Anat. Embryol. 210, 343–352.

Asami, T., Nakamura, R., Takaishi, M., Yoshida, H., Yoshimi, A., Whitford, T.J., Hirayasu, Y., 2018. Smaller volumes in the lateral and basal nuclei of the amygdala in patients with panic disorder. PloS One 13, e0207163.

Backhausen, L.L., Herting, M.M., Buse, J., Roessner, V., Smolka, M.N., Vetter, N.C., 2016. Quality control of structural MRI images applied using FreeSurfer—a hands-on workflow to rate motion artifacts. Front. Neurosci. 10, 558.

Babaev, O., Chatain, C.P., Krueger-Burg, D., 2018. Inhibition in the amygdala anxiety circuitry. Exp. Mol. Med. 50, 1–16.

Bang, J., Spina, S., Miller, B.L., 2015. Frontotemporal dementia. Lancet 386, 1672–1682.

Barnes, J., Bartlett, J.W., van de Pol, Laura, A., Loy, C.T., Scahill, R.I., Frost, C., Thompson, P., Fox, N.C., 2009. A meta-analysis of hippocampal atrophy rates in alzheimer's disease. Neurobiol. Aging 30, 1711–1723.

Bartsch, J.C., Schott, B.H., Behr, J., 2019. Hippocampal dysfunction in schizophrenia and aberrant hippocampal synaptic plasticity in rodent model psychosis: a selective review. Pharmacopsychiatry 2019. https://doi.org/10.1055/a-0960-9846.

Benarroch, E.E., 2015. The amygdala: functional organization and involvement in neurologic disorders. Neurology 84, 313–324.

Benson, B.E., Willis, M.W., Ketter, T.A., Speer, A., Kimbrell, T.A., Herscovitch, P., George, M.S., Post, R.M., 2014. Differential abnormalities of functional connectivity of the amygdala and hippocampus in unipolar and bipolar affective disorders. J. Affect. Disord. 168, 243–253.

Bouchard, T.P., Malykhin, N., Martin, W.W., Hanstock, C.C., Emery, D.J., Fisher, N.J., Camicioli, R.M., 2008. Age and dementia-associated atrophy predominates in the hippocampal head and amygdala in Parkinson's disease. Neurobiol. Aging 29, 1027–1039.

Braak, H., Braak, E., 1983. Neuronal types in the basolateral amygdaloid nuclei of man. Brain Res. Bull. 11, 349–365.

Brown, E.M., Pierce, M.E., Clark, D.C., Fischl, B.R., Iglesias, J.E., Milberg, W.P., McGlinchey, R.E., Salat, D.H., 2020. Test-retest reliability of FreeSurfer automated hippocampal subfield segmentation within and across scanners. Neuroimage 210, 116563.

Cao, B., Passos, I.C., Mwangi, B., Amaral-Silva, H., Tannous, J., Wu, M., Zunta-Soares, G.B., Soares, J.C., 2017. Hippocampal subfield volumes in mood disorders. Mol. Psychiatr. 22, 1352.

Catani, M., Dell'Acqua, F., De Schotten, M.T., 2013. A revised limbic system model for memory, emotion and behaviour. Neurosci. Biobehav. Rev. 37, 1724–1737.

Cavedo, E., Boccardi, M., Ganzola, R., Canu, E., Beltramello, A., Caltagirone, C., Thompson, P.M., Frisoni, G.B., 2011. Local amygdala structural differences with 3T MRI in patients with alzheimer disease. Neurology 76, 727–733.
Cullen, K.R., Westlund, M.K., Klimes-Dougan, B., Mueller, B.A., Houri, A., Eberly, L.E., Lim, K.O., 2014. Abnormal amygdala resting-state functional connectivity in adolescent depression. JAMA Psychiatry 71, 1138–1147.
Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage 9, 179–194.
Despotović, I., Goossens, B., Philips, W., 2015. MRI segmentation of the human brain: challenges, methods, and applications. Comput. Math. Methods Med.
Dewey, J., Hana, G., Russell, T., Price, J., McCaffrey, D., Harezlak, J., Sem, E., Anyanwu, J.C., Guttmann, C.R., Navia, B., 2010. Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. Neuroimage 51, 1334–1344.
Ding, S., 2013. Comparative anatomy of the prosubiculum, subiculum, presubiculum, postsubiculum, and parasubiculum in human, monkey, and rodent. J. Comp. Neurol. 521, 4145–4162.
Ding, S., Van Hoesen, G.W., 2015. Organization and detailed parcellation of human hippocampal head and body regions based on a combined analysis of cyto-and chemoarchitecture. J. Comp. Neurol. 523, 2233–2253.
Dupont, W.D., Plummer Jr., W.D., 1990. Power and sample size calculations: a review and computer program. Control. Clin. Trials 11, 116–128.
Dupont, W.D., Plummer Jr., W.D., 1998. Power and sample size calculations for studies involving linear regression. Contr. Clin. Trials 19, 589–601.
Driessen, M., Herrmann, J., Stahl, K., Zwaan, M., Meier, S., Hill, A., Osterheider, M., Petersen, D., 2000. Magnetic resonance imaging volumes of the hippocampus and the amygdala in women with borderline personality disorder and early traumatization. Arch. Gen. Psychiatr. 57, 1115–1122.
Feher, J., 2017. 4.6 - the Chemical Senses, pp. 427–439.
Fischl, B., 2012. FreeSurfer. Neuroimage, vol. 62, pp. 774–781.
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis II: inflation, flattening, and surface-based coordinate system. Neuroimage 9, 195–207.
Frankó, E., Joly, O., Alzheimer's Disease Neuroimaging Initiative, 2013. Evaluating Alzheimer's disease progression using rate of regional hippocampal atrophy. PloS One 8, e71354.
Frisoni, G.B., Ganzola, R., Canu, E., Rüb, U., Pizzini, F.B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C., Thompson, P.M., 2008. Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. Brain 131, 3266–3276.
Ganzola, R., Maziade, M., Duchesne, S., 2014. Hippocampus and amygdala volumes in children and young adults at high-risk of schizophrenia: research synthesis. Schizophr. Res. 156, 76–86.
Grimm, O., Pohlack, S., Cacciaglia, R., Winkelmann, T., Plichta, M.M., Demirakca, T., Flor, H., 2015. Amygdalar and hippocampal volume: a comparison between manual segmentation, freesurfer and VBM. J. Neurosci. Methods 253, 254–261.
Gryglewski, G., Baldinger-Melich, P., Seiger, R., Godbersen, G.M., Michenthaler, P., Klöbl, M., Spurny, B., Kautzky, A., Vanicek, T., Kasper, S., 2019. Structural changes in amygdala nuclei, hippocampal subfields and cortical thickness following electroconvulsive therapy in treatment-resistant depression: longitudinal analysis. Br. J. Psychiatr. 214, 159–167.
Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 32, 180–194.
Heimer, L., De Olmos, J., Alheid, G., Pearson, J., Sakamoto, N., Shinoda, K., Marksteiner, J., Switzer, R., 1999. The Human Basal Forebrain Part II the Primate Nervous System, Part III.
Helms, G., 2016. Segmentation of human brain using structural MRI. Magnetic Reson. Mater. Phys. Biol. Med. 29, 111–124.
Howard, M.W., Eichenbaum, H., 2015. Time and space in the hippocampus. Brain Res. 1621, 345–354.
Iglesias, J.E., Augustinack, J.C., Nguyen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A.C., Wald, L.L., 2015. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. Neuroimage 115, 117–137.
Iglesias, J.E., Van Leemput, K., Augustinack, J., Insausti, R., Fischl, B., Reuter, M., 2016. Alzheimer's Disease Neuroimaging Initiative, 2016. Bayesian longitudinal segmentation of hippocampal substructures in brain MRI using subject-specific atlases. Neuroimage 141, 542–555.
Iscan, Z., Jin, T.B., Kendrick, A., Szeglin, B., Lu, H., Trivedi, M., Fava, M., McGrath, P.J., Weissman, M., Kurian, B.T., 2015. Test–retest reliability of FreeSurfer measurements within and between sites: effects of visual approval process. Hum. Brain Mapp. 36, 3472–3485.
Janak, P.H., Tye, K.M., 2015. From circuits to behaviour in the amygdala. Nature 517, 284.
Janiri, D., Sani, G., Rossi, P.D., Piras, F., Iorio, M., Banaj, N., Giuseppin, G., Spinazzola, E., Maggiora, M., Ambrosi, E., 2017. Amygdala and hippocampus volumes are differently affected by childhood trauma in patients with bipolar disorders and healthy controls. Bipolar Disord. 19, 353–362.
Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., Fischl, B., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 46, 177–192.
Jovicich, J., Marizzoni, M., Sala-Llonch, R., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., Wiltfang, J., Roccatagliata, L., Nobili, F., 2013. Brain morphometry reproducibility in multi-center 3 T MRI studies: a comparison of cross-sectional and longitudinal segmentations. Neuroimage 83, 472–484.
Jovicich, J., Marizzoni, M., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., Wiltfang, J., Roccatagliata, L., Picco, A., Nobili, F., 2014. Multisite longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging of healthy elderly subjects. Neuroimage 101, 390–403.
Kemppainen, S., Jolkkonen, E., Pitkänen, A., 2002. Projections from the posterior cortical nucleus of the amygdala to the hippocampal formation and parahippocampal region in rat. Hippocampus 12, 735–755.
Kim, J., Pignatelli, M., Xu, S., Itohara, S., Tonegawa, S., 2016. Antagonistic negative and positive neurons of the basolateral amygdala. Nat. Neurosci. 19, 1636.
Knierim, J.J., 2015. The hippocampus. Curr. Biol. 25, R1116–R1121.
Konrad, C., Ukas, T., Nebel, C., Arolt, V., Toga, A.W., Narr, K.L., 2009. Defining the human hippocampus in cerebral magnetic resonance imaging- an overview of current segmentation protocols. Neuroimage 47, 1185–1195.
Krabbe, S., Gründemann, J., Lüthi, A., 2018. Amygdala inhibitory circuits regulate associative fear conditioning. Biol. Psychiatr. 83, 800–809.
Leal, S.L., Noche, J.A., Murray, E.A., Yassa, M.A., 2017. Disruption of amygdala–entorhinal–hippocampal network in late-life depression. Hippocampus 27, 464–476.
Marizzoni, M., Antelmi, L., Bosch, B., Bartrés-Faz, D., Müller, B.W., Wiltfang, J., Fiedler, U., Roccatagliata, L., Picco, A., Nobili, F., 2015. Longitudinal reproducibility of automatically segmented hippocampal subfields: A multisite E uropean 3T study on healthy elderly. Hum. Brain Mapp. 36, 3516–3527.
Montagrin, A., Saiote, C., Schiller, D., 2018. The social hippocampus. Hippocampus 28, 672–679.
Morey, R.A., Selgrade, E.S., Wagner, H.R., Huettel, S.A., Wang, L., McCarthy, G., 2010. Scan–rescan reliability of subcortical brain volumes derived from automated segmentation. Hum. Brain Mapp. 31, 1751–1762.
Morey, R.A., Petty, C.M., Xu, Y., Hayes, J.P., Wagner II, H.R., Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G., 2009. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. Neuroimage 45, 855–866.
Mueller, S.G., Yushkevich, P.A., Das, S., Wang, L., Van Leemput, K., Iglesias, J.E., Alpert, K., Mezher, A., Ng, P., Paz, K., 2018. Systematic comparison of different techniques to measure hippocampal subfield volumes in ADNI2. Neuroimage: Clinic 17, 1006–1018.
Mulder, E.R., de Jong, R.A., Knol, D.L., van Schijndel, R.A., Cover, K.S., Visser, P.J., Barkhof, F., Vrenken, H., Alzheimer's Disease Neuroimaging Initiative, 2014. Hippocampal volume change measurement: quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. Neuroimage 92, 169–181.
Murray, R.J., Brosch, T., Sander, D., 2014. The functional profile of the human amygdala in affective processing: insights from intracranial recordings. Cortex 60, 10–33.
Pipitone, J., Park, M.T.M., Winterburn, J., Lett, T.A., Lerch, J.P., Pruessner, J.C., Lepage, M., Voineskos, A.N., Chakravarty, M.M., Alzheimer's Disease Neuroimaging Initiative, 2014. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. Neuroimage 101, 494–512.
Poppenk, J., Evensmoen, H.R., Moscovitch, M., Nadel, L., 2013. Long-axis specialization of the human hippocampus. Trends Cognit. Sci. 17, 230–240.
Prestia, A., Cavedo, E., Boccardi, M., Muscio, C., Adorni, A., Geroldi, C., Bonetti, M., Thompson, P.M., Frisoni, G.B., 2015. Hippocampal and amygdalar local structural differences in elderly patients with schizophrenia. Am. J. Geriatr. Psychiatr. 23, 47–58.
Prestia, A., Boccardi, M., Galluzzi, S., Cavedo, E., Adorni, A., Soricelli, A., Bonetti, M., Geroldi, C., Giannakopoulos, P., Thompson, P., 2011. Hippocampal and amygdalar volume changes in elderly patients with alzheimer's disease and schizophrenia. Psychiatr. Res. Neuroimaging 192, 77–83.
Rajaratnam, N., 1960. Reliability formulas for independent decision data when reliability data are matched. Psychometrika 25, 261–271.
Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: a robust approach. Neuroimage 53, 1181–1196.
Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–1418.
Rolls, E.T., 2015. Limbic systems for emotion and for memory, but no single limbic system. Cortex 62, 119–157.
Rich, A.M., Cho, Y.T., Tang, Y., Savic, A., Krystal, J.H., Wang, F., Xu, K., Anticevic, A., 2016. Amygdala volume is reduced in early course schizophrenia. Psychiatr. Res. Neuroimaging 250, 50–60.
Rossi, R., Lanfredi, M., Pievani, M., Boccardi, M., Beneduce, R., Rillosi, L., Giannakopoulos, P., Thompson, P.M., Rossi, G., Frisoni, G.B., 2012. Volumetric and topographic differences in hippocampal subdivisions in borderline personality and bipolar disorders. Psychiatr. Res. Neuroimaging 203, 132–138.
Sah, P., Faber, E.L., Lopez de Armentia, M., Power, J., 2003. The amygdaloid complex: anatomy and physiology. Physiol. Rev. 83, 803–834.

Saygin, Z., Kliemann, D., Iglesias, J., van der Kouwe, André, J.W., Boyd, E., Reuter, M., Stevens, A., Van Leemput, K., McKee, A., Frosch, M.P., 2017. High-resolution magnetic resonance imaging reveals nuclei of the human amygdala: manual segmentation to automatic atlas. Neuroimage 155, 370–382.
Schmahl, C., Berne, K., Krause, A., Kleindienst, N., Valerius, G., Vermetten, E., Bohus, M., 2009. Hippocampus and amygdala volumes in patients with borderline personality disorder with or without posttraumatic stress disorder. J. Psychiatry Neurosci. 34, 289–295.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420.
Strange, B.A., Witter, M.P., Lein, E.S., Moser, E.I., 2014. Functional organization of the hippocampal longitudinal axis. Nat. Rev. Neurosci. 15, 655–669.
Tae, W.S., Kim, S.S., Lee, K.U., Nam, E., Kim, K.W., 2008. Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. Neuroradiology 50, 569.
Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imag. 15, 29.
Tyszka, J.M., Pauli, W.M., 2016. In vivo delineation of subdivisions of the human amygdaloid complex in a high-resolution group template. Hum. Brain Mapp. 37, 3979–3998.
Ubeda-Bañon, I., Novejarque, A., Mohedano-Moriano, A., Pro-Sistiaga, P., de la Rosa-Prieto, C., Insausti, R., Martinez-Garcia, F., Lanuza, E., Martinez-Marcos, A., 2007. Projections from the posterolateral olfactory amygdala to the ventral striatum: neural basis for reinforcing properties of chemical stimuli. BMC Neurosci. 8, 103.
van den Burg, E.H., Stoop, R., 2019. Neuropeptide signalling in the central nucleus of the amygdala. Cell Tissue Res. 375, 93–101.
Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L.L., Augustinack, J., Dickerson, B.C., Golland, P., Fischl, B., 2009. Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. Hippocampus 19, 549–557.
Viviani, R., Pracht, E.D., Brenner, D., Beschoner, P., Stingl, J.C., Stöcker, T., 2017. Multimodal MEMPRAGE, FLAIR, and R2* segmentation to resolve dura and vessels from cortical gray matter. Front. Neurosci. 11, 258.
Wassum, K.M., Izquierdo, A., 2015. The basolateral amygdala in reward learning and addiction. Neurosci. Biobehav. Rev. 57, 271–283.
Weniger, G., Lange, C., Sachsse, U., Irle, E., 2009. Reduced amygdala and hippocampus size in trauma-exposed women with borderline personality disorder and without posttraumatic stress disorder. J. Psychiatry Neurosci. 34, 383–388.
Wijeratne, C., Sachdev, S., Wen, W., Piguet, O., Lipnicki, D.M., Malhi, G.S., Mitchell, P.B., Sachdev, P.S., 2013. Hippocampal and amygdala volumes in an older bipolar disorder sample. Int. Psychogeriatr. 25, 54–60.
Wisse, L.E., Biessels, G.J., Geerlings, M.I., 2014. A critical appraisal of the hippocampal subfield segmentation package in FreeSurfer. Front. Aging Neurosci. 6, 261.
Wonderlick, J.S., Ziegler, D.A., Hosseini-Varnamkhasti, P., Locascio, J.J., Bakkour, A., van der Kouwe, A., Triantafyllou, C., Corkin, S., Dickerson, B.C., 2009. Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. Neuroimage 44, 1324–1333.
Worker, A., Dima, D., Combes, A., Crum, W.R., Streffer, J., Einstein, S., Mehta, M.A., Barker, G.J., Cr Williams, S., O'daly, O., 2018. Test–retest reliability and longitudinal analysis of automated hippocampal subregion volumes in healthy ageing and A lzheimer's disease populations. Hum. Brain Mapp. 39, 1743–1754.
Yang, Y., Wang, J., 2017. From structure to behavior in basolateral amygdala-hippocampus circuits. Front. Neural Circ. 11, 86.
Yushkevich, P.A., Amaral, R.S., Augustinack, J.C., Bender, A.R., Bernstein, J.D., Boccardi, M., Bocchetta, M., Burggren, A.C., Carr, V.A., Chakravarty, M.M., 2015. Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. Neuroimage 111, 526–541.
Zarei, M., Beckmann, C.F., Binnewijzend, M.A., Schoonheim, M.M., Oghabian, M.A., Sanz-Arigita, E.J., Scheltens, P., Matthews, P.M., Barkhof, F., 2013. Functional segmentation of the hippocampus in the healthy human brain and in alzheimer's disease. Neuroimage 66, 28–35.
Zhong, Q., Xu, H., Qin, J., Zeng, L., Hu, D., Shen, H., 2019. Functional parcellation of the hippocampus from resting-state dynamic functional connectivity. Brain Res. 1715, 165–175.
Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C., 1994. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans. Med. Imag. 13, 716–724.
Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells III, W.M., Jolesz, F.A., Kikinis, R., 2004. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Acad. Radiol. 11, 178–189.
Zuo, X., Anderson, J.S., Bellec, P., Birn, R.M., Biswal, B.B., Blautzik, J., Breitner, J.C., Buckner, R.L., Calhoun, V.D., Castellanos, F.X., 2014. An open science resource for establishing reliability and reproducibility in functional connectomics, 1, 140049. Scientific Data.